A-OKVQA: A Benchmark for Visual Question Answering using World Knowledge

Year: 2022 | Citations: 729 | Authors: Dustin Schwenk, Apoorv Khandelwal, Christopher Clark, Kenneth Marino, Roozbeh Motta

---

Abstract

The Visual Question Answering (VQA) task aspires to provide a meaningful testbed for the development of AI models that can jointly reason over visual and natural language inputs. Despite a proliferation of VQA datasets, this goal is hindered by a set of common limitations. These include a reliance on relatively simplistic questions that are repetitive in both concepts and linguistic structure, little world knowledge needed outside of the paired image, and limited reasoning required to arrive at the correct answer. We introduce A-OKVQA, a crowdsourced dataset composed of a diverse set of about 25K questions requiring a broad base of commonsense and world knowledge to answer. In contrast to the existing knowledge-based VQA datasets, the questions generally cannot be answered by simply querying a knowledge base, and instead require some form of commonsense reasoning about the scene depicted in the image. We demonstrate the potential of this new dataset through a detailed analysis of its contents and baseline performance measurements over a variety of state-of-the-art vision-language models. Project page: http://a-okvqa.allenai.org/